On the estimation and correction of discretization error in molecular dynamics averages

Nana Arizumi[a,*], Stephen D. Bond[b]

[a] *Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States*
[b] *Multiphysics Simulation Technologies Department, Sandia National Laboratories, Albuquerque, NM 87185, United States*

Abstract

The computation of statistical averages is one of the most important applications of molecular dynamics simulation, allowing for the estimation of macroscopic physical quantities through averages of observables sampled along microscopic trajectories. In this article, we investigate the impact of discretization error on the accuracy of molecular dynamics averages. Given a Hamiltonian system and a symplectic integrator, new weighting methods are derived to better approximate averages of certain observables, without changing the system or integrator. These new methods are shown to reduce discretization error and enhance the order of accuracy without high-overhead calculations.

Keywords: molecular dynamics, statistical mechanics, ensemble averages, symplectic integrators, Nosé-Poincaré Hamiltonian

1. Introduction

Classical molecular dynamics (MD) is a widely used tool in a variety of fields, including computational chemistry, physics, and materials science [3, 14, 29]. Statistical mechanics provides a connection between macroscopic physical properties and ensemble averages of microscopic molecular motion [22, 32], which can be simulated numerically using MD. However, accurate simulation of physically relevant phenomena is a challenging task, requiring extremely long simulations on large computers due to the wide range of spatial and temporal scales, which imposes severe limitations on the timestep.

The primary goal of MD is to compute averages and correlations along a trajectory, rather than the trajectory itself. For this reason, it is important to use theory and methods which focus on the statistical properties of the trajectory. Although computed numerical trajectories may fluctuate and drift away from the exact dynamics of the molecular model, the resulting averages remain valid provided the numerical trajectories have the correct statistical properties.

Techniques for estimating and correcting statistical and discretization error in MD simulation has been explored before [7, 8, 10, 11, 13, 17, 26, 30, 33]. The foundation of our analysis of MD averages is based on backward error analysis for symplectic integrators applied to Hamiltonian systems. In section 2, we introduce specific Hamiltonian systems and statistical ensembles used in MD simulation. Due to the chaotic nature of MD trajectories, small numerical errors lead to large forward error, and as a consequence, forward error analysis is not applicable for understanding MD averages. Instead, our focus will be on backward error analysis, using the method of modified equations. This involves deriving a modified vector field, the dynamics of which is arbitrarily close to the numerical trajectory [4, 15–17, 23, 26–28, 31]. Under certain assumptions of ergodicity, the difference between exact and numerical averages can be analyzed by measuring the change in the average after perturbation to the vector field. In section 3, we provide a brief review of backward error analysis, focusing on the application of symplectic integrators to Hamiltonian systems, which has considerable advantages since the modified equations are also Hamiltonian.

*Corresponding author.
Email addresses: arizumi@illinois.edu (Nana Arizumi), sdbond@sandia.gov (Stephen D. Bond)

The error in MD averages can be roughly decomposed into two terms: statistical error and discretization error. Statistical error can be attributed to the use of finite-length trajectories, which results in insufficient sampling. On the other hand, the discretization error is associated with the large timestep size. The focus of this article is on the discretization error, which under certain assumptions, can be interpreted as a modification to the Hamiltonian. In section 4, we will introduce methods to approximate this error for two systems. This first system is classical Hamiltonian dynamics, corresponding to the microcanonical (constant energy) ensemble. For this system, we introduce techniques for measuring and correcting the discretization error in numerical averages. The second system is the extended Nosé Hamiltonian system used to generate dynamics from the canonical (constant temperature) ensemble. For this system, we derive a formulation for determining a "weighting" factor introduced in [5], which again allows for the correction of discretization error in numerical averages. These new methods introduce additional computational overhead during the post-processing of simulation data. In section 5, we derive finite difference estimates which dramatically reduce the complexity. Finally, we present some numerical results in section 6.

2. Molecular dynamics averages

One of the most important applications of classical MD is the computation of a statistical average along a trajectory, or collection of trajectories [3, 5, 14, 18]. In this paper, we will assume that each trajectory is generated by solving an ordinary differential equation (ODE) of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{z}\left(t\right) = F\left(\mathbf{z}\left(t\right)\right),$$

with $\mathbf{z}\left(0\right) = \mathbf{z}_0 \in \mathbb{R}^{6N}$. Here, $\mathbf{z}\left(t\right)^T = \left[\mathbf{q}\left(t\right)^T \mathbf{p}\left(t\right)^T\right]$ is a vector containing the positions, $\mathbf{q}\left(t\right) \in \mathbb{R}^{3N}$, and momenta, $\mathbf{p}\left(t\right) \in \mathbb{R}^{3N}$, of N atoms at time t. For an isolated system, the system of equations for Hamiltonian $H : \mathbb{R}^{6N} \to \mathbb{R}$ is

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{z}\left(t\right) = J\nabla_{\mathbf{z}} H\left(\mathbf{z}\left(t\right)\right),$$

where J is the skew-symmetric symplectic structure matrix

$$J := \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

For the special case of Newton's equations of motion,

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = m_i^{-1} p_i, \qquad \frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial U(q)}{\partial q_i},$$

where U is the potential energy function and m_i is the mass associated with the ith component of the momenta, one can show that the Hamiltonian is

$$H\left(\mathbf{q}, \mathbf{p}\right) = \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p} + U\left(\mathbf{q}\right),$$

where M is a diagonal mass matrix.

For a given observable $A : \mathbb{R}^{6N} \to \mathbb{R}$, the time average along a trajectory with initial condition $\mathbf{z}_0$ is defined as

$$\left\langle A\right\rangle_{\mathbf{z}_0,\mathrm{time}} := \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau A\left(\mathbf{z}\left(t\right)\right)\mathrm{d}t.$$

On the other hand, the ensemble average [9, 18, 22, 32] is given by

$$\left\langle A\right\rangle_{\mathrm{ens}} = \int_\Omega A\left(\mathbf{z}\right)\rho_{\mathrm{ens}}\left(\mathbf{z}\right)\mathrm{d}\mathbf{z},$$

where ρ_{ens} is an invariant distribution for the dynamics. If the system is ergodic, one can show that these averages are the same,

$$\langle A \rangle_{\text{ens}} = \langle A \rangle_{\text{time}},$$

for all initial conditions outside a set of measure zero [32].

The evolution of a distribution of configurations in phase space, ρ, under the flow of a vector field, F, can be described by the Liouville equation [32]

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho F) = 0.$$

Alternatively, this can be rewritten in terms of a material derivative, which results in

$$\frac{D\rho}{Dt} + \rho \nabla \cdot F = 0.$$

If the vector field is Hamiltonian, one can show that it is divergence-free, the material derivative of ρ is zero, and

$$\rho \propto \delta[H - E]$$

is invariant, where H is the Hamiltonian and E is the energy.

For a system in contact with a heat bath (at constant temperature), the configurations are distributed according to the Gibbs (or canonical) ensemble,

$$\rho_c(\mathbf{z}) = \frac{1}{C_c} \exp(-\frac{1}{k_B T} H(\mathbf{z})). \tag{1}$$

Here, k_B is Boltzmann's constant, T is the temperature, and C_c is a normalizing constant.

To show that an invariant distribution is unique, one must prove the system is ergodic. Due to the presence of first integrals (e.g., conservation of energy or momentum) a trajectory may be confined to a submanifold of phase space. One must account for these first integrals to make a connection between the dynamics of a single trajectory (or small set of trajectories) and the invariant distribution. For example, for the microcanonical ensemble, energy is conserved, which means the Hamiltonian is a first integral of the motion. If the system also has in addition k first integrals I_i, with constant values C_i, then the invariant distribution ρ becomes

$$\rho = \frac{1}{C} \delta[H - E] \delta[I_1 - C_1] \cdots \delta[I_k - C_k],$$

where E is the energy and C is a normalizing constant.

3. Backward error analysis

In traditional forward error analysis, the error in a trajectory generated by an rth-order numerical method is bounded by

$$\|\mathbf{z}(t_n) - \tilde{\mathbf{z}}_n\| \leq C_1 \Delta t^r \exp(C_2 t_n),$$

where $\{\mathbf{z}(t) | t \geq 0\}$ and $\{\tilde{\mathbf{z}}_n | n = 0, 1, \ldots\}$ are the exact and numerical trajectories, respectively. Here, Δt is the timestep size, and the constants C_1 and C_2 depend on the particular vector field and the numerical method. For MD, C_2 is positive, thus a bound of this form is useful only over short time intervals. In addition, the ergodic hypothesis implies that statistical averages should not depend on the details of the initial conditions. Therefore, a different technique is needed to analyze the error in long time averages.

A more practical error bound for MD can be derived using backward error analysis. In this framework, discretization error arising from the application of a numerical integrator is interpreted as a perturbation to the vector field, using the method of modified equations. For each compact set K of the phase space, there exists a finite constant C_K such that

$$\left\| \tilde{F}(\mathbf{z}) - F(\mathbf{z}) \right\| \leq C_K \Delta t^r, \ \forall \mathbf{z} \in K,$$

where $F(\mathbf{z})$ is the vector field for the exact trajectory, $\tilde{F}(\mathbf{z})$ is the modified vector field, and r is the order of the numerical method used to generate the trajectory. For a symplectic integrator applied to a Hamiltonian system, one can show that there exists a modified system whose dynamics is exponentially close to the numerical trajectory for an exponentially long time [4, 15, 17, 21, 23, 26–28, 31].

Although the numerical trajectory does not exactly interpolate the flow of the modified vector field, it is useful as a truncated series in the context of geometric integrators. If the numerical method preserves certain geometric structures, then the modified equations also preserve the same structures. In particular, if we apply a symplectic integrator to a Hamiltonian vector field, the modified vector field is also Hamiltonian. Here, we denote this modified Hamiltonian as $\tilde{H}$. One can show that

$$\tilde{H}(\mathbf{z}) = H(\mathbf{z}) + \Delta t^r H_1(\mathbf{z}, \Delta t) + \mathcal{O}\left(\Delta t^{r+1}\right),$$

for small Δt, where H_1 is the first term in a series expansion.

4. Errors in averages

The error in a time average along a numerical trajectory can be expressed as a sum of two terms [7, 8],

$$|\langle A\rangle_{\text{exact}} - \langle A\rangle_{\text{num}}| \leq C_{\text{s}}/\sqrt{t} + C_{\text{t}}\Delta t^r.$$

Here, $\langle A\rangle_{\text{exact}}$ is the average of A over the exact trajectory using an infinite time interval and $\langle A\rangle_{\text{num}}$ is the average of A over a finite-length numerical trajectory. The first term is statistical error, which results from the finite length of the simulation. Typical statistical error converges at a rate inversely proportional to the square-root of the simulation time. The second term is truncation error, which results from the non-zero step size in the rth-order numerical method used to generate the trajectory. The goal of this paper is to save computation time by estimating and correcting for discretization error in computed averages using a single simulation, without modifying the numerical integrator, and without performing multiple simulations using differing timestep sizes.

4.1. Expansion of the delta function

In this section, we show how to calculate the change in an ensemble average resulting from a small perturbation to the Hamiltonian using an expansion of a delta function. We use notation $\langle A\rangle_H$ for an average of A taken from the Hamiltonian surface, $H^{-1}(E)$, and $\langle A\rangle_{\tilde{H}}$ for an average of A taken from the modified Hamiltonian surface, $\tilde{H}^{-1}(\tilde{E})$. A related method was introduced in [30]. Without loss of generality, we will assume that H and $\tilde{H}$ are shifted in such a way that $H(\mathbf{z}_0) = \tilde{H}(\mathbf{z}_0) = 0$, where $\mathbf{z}_0$ is the initial condition. The direction for the expansion is determined by the choice of a vector field $\mathbf{w}$, which leads to an expression for the change in the average characterized by the following proposition.

Proposition 1. *Let H be a C^1 energy function with $H^{-1}(0)$ compact. Let $\tilde{H}$ be a C^1 modified energy function, with $\tilde{H} = H + \epsilon G$ and $\tilde{H}^{-1}(0)$ compact. Suppose $\mathbf{w}$ is a given C^1 vector field with $\mathbf{w}\cdot\nabla H \neq 0$ everywhere on $H^{-1}(0)$. The change of the microcanonical average of a given C^1 observable A can be expressed as*

$$\langle A\rangle_H - \langle A\rangle_{\tilde{H}} = \langle\nabla\cdot(\mathbf{u}\,A)\rangle_{\tilde{H}} - \langle A\rangle_{\tilde{H}}\langle\nabla\cdot\mathbf{u}\rangle_{\tilde{H}} + \mathcal{O}\left(\epsilon^2\right),$$

where

$$\mathbf{u} := (\tilde{H} - H)\frac{\mathbf{w}}{\mathbf{w}\cdot\nabla H} = \epsilon\, G\frac{\mathbf{w}}{\mathbf{w}\cdot\nabla H},$$

assuming $\epsilon << 1$.

Proof. Expanding $\delta\big(H\big)$ about $\delta\big(\tilde{H}\big)$, we have

$$\int A\,\delta\big(H\big)\mathrm{d}z = \int A\,\delta\big(\tilde{H}\big)\mathrm{d}z - \epsilon\int G\,A\,\delta'\big(\tilde{H}\big)\mathrm{d}z + \mathcal{O}\left(\epsilon^2\right). \tag{2}$$

Using the directional derivative, it can be shown that

$$\int G\,A\,\delta'\big(\tilde{H}\big)\mathrm{d}z = \int G\,A\,\frac{\mathbf{w}}{\mathbf{w}\cdot\nabla\tilde{H}}\cdot\nabla\delta\big(\tilde{H}\big)\mathrm{d}z = -\int\nabla\cdot\left(G\,A\,\frac{\mathbf{w}}{\mathbf{w}\cdot\nabla\tilde{H}}\right)\delta\big(\tilde{H}\big)\mathrm{d}z$$

for any differentiable vector field $\mathbf{w}$. The boundary term vanishes under reasonable assumptions on $\tilde{H}$ at the boundary phase space. Inserting this expression into (2) and using the definition of $\mathbf{u}$, we have

$$\int A\delta\big(H\big)\mathrm{d}z = \int A\,\delta\big(\tilde{H}\big)\mathrm{d}z + \int\nabla\cdot(\mathbf{u}A)\,\delta\big(\tilde{H}\big)\mathrm{d}z + \mathcal{O}\left(\epsilon^2\right),$$

and hence

$$\langle A\rangle_H = \frac{\int A\delta\big(H\big)\mathrm{d}z}{\int\delta\big(H\big)\mathrm{d}z} = \frac{\langle A\rangle_{\tilde{H}} + \langle\nabla\cdot(\mathbf{u}A)\rangle_{\tilde{H}}}{1+\langle\nabla\cdot\mathbf{u}\rangle_{\tilde{H}}} + \mathcal{O}\left(\epsilon^2\right).$$

Subtracting $\langle A\rangle_{\tilde{H}}$ from both sides and combining terms, we have

$$\langle A\rangle_H - \langle A\rangle_{\tilde{H}} = \frac{\langle\nabla\cdot(\mathbf{u}A)\rangle_{\tilde{H}} - \langle A\rangle_{\tilde{H}}\langle\nabla\cdot\mathbf{u}\rangle_{\tilde{H}}}{1+\langle\nabla\cdot\mathbf{u}\rangle_{\tilde{H}}} + \mathcal{O}\left(\epsilon^2\right).$$

Note that since the numerator is of order ϵ and the denominator is 1 plus order ϵ, the denominator can be ignored, and we arrive at the desired result. □

There is some freedom in the choice of the vector field $\mathbf{w}$. In principle, almost any $\mathbf{w}$ will work, so long as the resulting $\mathbf{u}$ is defined and differentiable everywhere. A natural choice is to set $\mathbf{w} = \nabla H$, which reduces the possibility of division by zero.

One disadvantage of using the expression in Proposition 1 is that it requires differentiating the observable, A. This is typically inconvenient in the context of a molecular dynamics software package which supports a wide range of user supplied observables. To avoid explicit calculation of the derivative of A, we introduce the following mapping

$$\hat{\mathbf{f}}(\mathbf{z}) := \mathbf{z} + \frac{\tilde{H}(\mathbf{z}) - H(\mathbf{z})}{\nabla H(\mathbf{z})\cdot\mathbf{w}(\mathbf{z})}\mathbf{w}(\mathbf{z}) = \mathbf{z} + \mathbf{u}(\mathbf{z}),$$

where $\mathbf{w}$ and $\mathbf{u}$ are as defined in Proposition 1. This mapping can be derived by expanding $H(\hat{\mathbf{f}}(\mathbf{z})) = 0$ under the assumption that $\tilde{H}(\mathbf{z}) = 0$ and $H(\mathbf{z})$ is small. Composing A with $\hat{\mathbf{f}}$ and expanding, we find

$$A(\hat{\mathbf{f}}(\mathbf{z})) = A(\mathbf{z}) + \mathbf{u}(\mathbf{z})\cdot\nabla A(\mathbf{z}) + \mathcal{O}(\epsilon^2),$$

assuming $\mathbf{u} = \mathcal{O}(\epsilon)$. Using this expansion, we can approximate

$$\nabla\cdot(\mathbf{u}(\mathbf{z})A(\mathbf{z})) = A(\mathbf{z})\nabla\cdot\mathbf{u}(\mathbf{z}) + A(\hat{\mathbf{f}}(\mathbf{z})) - A(\mathbf{z}) + \mathcal{O}(\epsilon^2),$$

and the expression in Proposition 1 can be rewritten as

$$\langle A\rangle_H = \langle A\,\nabla\cdot\mathbf{u}\rangle_{\tilde{H}} + \langle A\circ\hat{\mathbf{f}}\rangle_{\tilde{H}} - \langle A\rangle_{\tilde{H}}\langle\nabla\cdot\mathbf{u}\rangle_{\tilde{H}} + \mathcal{O}\left(\epsilon^2\right).$$

For the special case when $\mathbf{w} = \nabla H$, we have

$$\langle A\rangle_H = \frac{\left\langle\omega\,A\circ\hat{\mathbf{f}}\right\rangle_{\tilde{H}}}{\langle\omega\rangle_{\tilde{H}}} + \mathcal{O}(\epsilon^2), \tag{3}$$

where the weight, ω, is defined as

$$\omega := \frac{\nabla\tilde{H}\cdot\nabla H}{\nabla H\cdot\nabla H} - H\frac{\nabla\cdot(\nabla H)}{\nabla H\cdot\nabla H} + 2H\frac{\nabla H^T H''\nabla H}{(\nabla H\cdot\nabla H)^2}. \tag{4}$$

Note that $\omega = 1 + \nabla \cdot u$ on $\tilde{H^{-1}}(0)$.

Alternatively, if we select $\mathbf{w}$ such that $\mathbf{w} \cdot \nabla A = 0$, then

$$\langle \nabla \cdot (\mathbf{u}\, A) \rangle_{\tilde{H}} = \langle A \nabla \cdot \mathbf{u} \rangle_{\tilde{H}},$$

and we do not need to compute the gradient of A. For example, if A is a function of only the position q, then we can select $\mathbf{w}\,(q,p)^T = \left[\mathbf{0}^T, \mathbf{v}(\mathbf{q},\mathbf{p})^T\right]$ for any vector field $\mathbf{v}$.

If the system has multiple invariants, Proposition 1 can be generalized for special cases. Consider a component of the linear or angular momentum, which is preserved by the Verlet method [17]. The average of A becomes

$$\langle A \rangle_H = \frac{1}{C} \int A \delta(I_1) \cdots \delta(I_k) \delta(H) dz,$$

where $I_1(z) = \cdots = I_k(z) = 0$ are the corresponding k-additional integral invariants. In general, $\mathbf{w}$ should be orthogonal to each ∇I_i. For example, we can choose

$$\mathbf{w} = \nabla H - \frac{\nabla H \cdot \nabla I_1}{\nabla I_1 \cdot \nabla I_1} \nabla I_1, \tag{5}$$

to orthogonalize $\mathbf{w}$ with respect to the first additional integral invariant preserved by the integrator, repeating the procedure for each additional invariant. This construction can be extended to first integrals not exactly preserved by the integrator if the corresponding modified first integrals are known. However, modified first integrals may not exist unless I is separable [34], i.e., $I = I_q(q) + I_p(p)$.

4.2. Constant temperature using the Nosé-Poincaré Hamiltonian

For the canonical ensemble case, we can use the Nosé-Poincaré Hamiltonian [6, 12]

$$H_{\mathrm{NP}} = s(H_{\mathrm{N}} - H_{\mathrm{N}}^0),$$

where H_{N}^0 is selected so that $H_{\mathrm{NP}}(\mathbf{z}_0) = 0$ for the initial condition $\mathbf{z}_0$, and H_{N} is the Nosé Hamiltonian [24, 25]

$$H_{\mathrm{N}} = H(\mathbf{q}, \tilde{\mathbf{p}}/s) + \frac{p_s^2}{2\mu} + g k_B T \ln s.$$

Here, H is the physical energy function, g is the total number of degrees of freedom of the physical system, k_B is the Boltzmann constant, T is the target temperature, μ is a thermostat mass parameter that effectively allows the strength of dynamic coupling to be adjusted, s is an extended position variable, p_s is an extended momentum variable, $\mathbf{q}$ is the vector of positions, and $\tilde{\mathbf{p}}$ is a vector of canonical momenta. The vector of canonical momenta, $\tilde{\mathbf{p}}$, is related to the physical momenta, $\mathbf{p}$, through a scaling $\mathbf{p} = \tilde{\mathbf{p}}/s$. Assuming ergodicity, one can show that

$$\langle A(\mathbf{q}, \tilde{\mathbf{p}}/s) \rangle_{H_{\mathrm{NP}}} = \langle A(\mathbf{q}, \mathbf{p}) \rangle_c,$$

where $\langle \cdot \rangle_{H_{\mathrm{NP}}}$ is the microcanonical average with respect to the Nosé-Poincaré Hamiltonian in extended phase space, and $\langle \cdot \rangle_c$ is the canonical average with respect to H,

$$\langle A(\mathbf{q}, \mathbf{p}) \rangle_c = \int A(\mathbf{q}, \mathbf{p}) \rho_c(\mathbf{q}, \mathbf{p})\, \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p},$$

where ρ_c is the canonical ensemble defined in (1). A proof of this result can be found in [6].

To derive a procedure for correcting averages, we assume that there exists a modified probability density function, $\tilde{\rho}$, which is close to the exact probability density function, ρ_c, and can be written in the form

$$\tilde{\rho}\,(\mathbf{z})\, \omega_c\,(\mathbf{z}) = \rho_{c,\mathrm{exact}}\,(\mathbf{z})\,,$$

where the spatially varying weighting factor, ω_c, is computed from backward error analysis [5]. Under certain assumptions, the weighting factor is nearly unity, $\omega_c\,(\mathbf{z}) = 1 + \mathcal{O}\,(\Delta t^r)$, and can be used to

post-process statistical averages of an observable A. Assuming a sufficiently long simulation, a higher-order approximation for the average of A can be computed using a weighted average along the numerical trajectory,

$$\langle A \rangle_{\text{exact}} = \langle A\, \omega_c \rangle_{\text{num}} / \langle \omega_c \rangle_{\text{num}} + \mathcal{O}\left(\Delta t^p\right),$$

where $p > r$. This is in contrast to the unweighted average which is only provides an rth-order approximation to the average.

Applying the generalized leapfrog algorithm (GLA) to the Nosé-Poincaré equations results in the following numerical algorithm

$$\left\{\begin{array}{rcl}
\tilde{\mathbf{p}}^{n+\frac{1}{2}} &=& \tilde{\mathbf{p}}^{n} - \dfrac{\Delta t}{2} s^n \nabla U(\mathbf{q}^n) \\
p_s^{n+\frac{1}{2}} &=& p_s^n + \dfrac{\Delta t}{2}\left(\dfrac{(\tilde{\mathbf{p}}^{n+\frac{1}{2}})^T \mathbf{M}^{-1} \tilde{\mathbf{p}}^{n+\frac{1}{2}}}{(s^n)^2} - g k_B T - \Delta H_{\text{N}}(\mathbf{q}^n, s^n, \tilde{\mathbf{p}}^{n+\frac{1}{2}}, p_s^{n+\frac{1}{2}}) \right) \\
s^{n+1} &=& s^n + \dfrac{\Delta t}{2}(s^n + s^{n+1}) p_s^{n+\frac{1}{2}}/\mu \\
\mathbf{q}^{n+1} &=& \mathbf{q}^n + \dfrac{\Delta t}{2}\left(\dfrac{1}{s^n} + \dfrac{1}{s^{n+1}} \right) \mathbf{M}^{-1} \tilde{\mathbf{p}}^{n+\frac{1}{2}} \\
p_s^{n+1} &=& p_s^{n+\frac{1}{2}} + \dfrac{\Delta t}{2}\left(\dfrac{(\tilde{\mathbf{p}}^{n+\frac{1}{2}})^T \mathbf{M}^{-1} \tilde{\mathbf{p}}^{n+\frac{1}{2}}}{(s^n)^2} - g k_B T - \Delta H_{\text{N}}(\mathbf{q}^{n+1}, s^{n+1}, \tilde{\mathbf{p}}^{n+\frac{1}{2}}, p_s^{n+\frac{1}{2}}) \right) \\
\tilde{\mathbf{p}}^{n+1} &=& \tilde{\mathbf{p}}^{n+\frac{1}{2}} - \dfrac{\Delta t}{2} s^{n+1} \nabla U(\mathbf{q}^{n+1})
\end{array}\right. , \tag{6}$$

where $\Delta H_N := H_N - H_N^0$, is the deviation in the Nosé Hamiltonian. This method, also known as the Nosé-Poincaré method [6], is symplectic, time-reversible, and second-order accurate.

In the following proposition, the modified distribution $\tilde{\rho}$ for the Nosé-Poincaré method is derived using the microcanonical distribution for the Nosé-Poincaré Hamiltonian over the extended phase space (including both the physical and the thermostat degrees of freedom). A weighting factor, ω_c, was previously derived for this method under the assumption that the constant H_{N}^0 is perturbed so that the modified Nosé-Poincaré Hamiltonian is a Poincaré transformation of a modified Nosé Hamiltonian, rather than setting $H_{\text{NP}}(\mathbf{z}) = 0$ (see [5] for details). Here, we follow the derivation presented in [5], but generalize the result to the case where the modified Nosé-Poincaré Hamiltonian is small, but not shifted to zero.

Proposition 2. *Assuming ergodicity, and that $g k_B T e^{\hat{\eta}} >> \mathcal{O}(H_{\text{N}}^0)$, a modified invariant distribution for Nosé-Poincaré method can be calculated as*

$$\tilde{\rho} = \frac{\rho_c}{\tilde{C}} \exp\{f_1(\mathbf{q},\mathbf{p})\} \left(1 + \sqrt{1 - \frac{1}{g}} \frac{H_{\text{N}}^0}{k_B T} \exp\left\{ \frac{-f_1(\mathbf{q},\mathbf{p})}{g} - \frac{H - H_{\text{N}}^0}{g k_B T} \right\} \right) + \mathcal{O}(\Delta t^4), \tag{7}$$

where $\tilde{C}$ is a constant,

$$f_1(\mathbf{q},\mathbf{p}) = \frac{-(\Delta t)^2}{24 k_B T} \left(\sum_j \sum_k \frac{2 p_j p_k U_{q_j q_k}}{m_j m_k} - \sum_j \frac{U_{q_j}^2}{m_j} - \frac{1}{\mu} \left(\sum_j \frac{p_j^2}{m_j} - g k_B T \right)^2 \right),$$

with timestep Δt, and

$$\rho_c = \frac{1}{C_c} \exp\{ -\frac{1}{k_B T} H(\mathbf{q},\mathbf{p}) \}.$$

Proof. Following the derivation in [5] equation (3.9), the modified distribution for the modified Hamiltonian is derived as

$$\tilde{\rho}(\mathbf{q},\mathbf{p})\, d\mathbf{p}\, d\mathbf{q} = \frac{1}{C} \int\int \delta\left[s\left(H_{\text{N}} - H_{\text{N}}^0 + \epsilon H_1 - H_{\text{N}}^0/s \right) \right] d\tilde{p}\, dq\, dp_s\, ds.$$

Applying a change of variables, $\tilde{\mathbf{p}}/s$ to $\mathbf{p}$ and $\ln s$ to η, gives

$$\tilde{\rho} = \frac{1}{C} \int\int e^{(N_f+1)\eta} \delta\left[e^{\eta} \left(H(\mathbf{q},\mathbf{p}) + p_s^2/2\mu + g k_B T \eta + \epsilon H_1(\mathbf{q}, e^{\eta}, \mathbf{p}, p_s) - H_{\text{N}}^0 - H_{\text{N}}^0/e^{\eta} \right) \right] dp_s\, d\eta.$$

Denote $r(\eta) = e^{\eta}\left(H(\mathbf{q},\mathbf{p}) + \frac{p_s^2}{2\mu} + gk_BT\eta + \epsilon H_1(\mathbf{q},e^{\eta},\mathbf{p},p_s) - H_{\mathrm{N}}^0 - \frac{H_{\mathrm{N}}^0}{e^{\eta}}\right)$ and the root of $r(\eta)$ as $\hat{\eta}$. Using the identity $\delta[r(\eta)] = \delta[\eta-\hat{\eta}]/|r'(\hat{\eta})|$,

$$\tilde{\rho} = \frac{1}{C}\int e^{N_f\hat{\eta}}\left|gk_BT + \epsilon\frac{\partial}{\partial\eta}H_1(\mathbf{q},e^{\eta},\mathbf{p},p_s) + \frac{H_{\mathrm{N}}^0}{e^{\eta}}\right|_{\eta=\hat{\eta}}^{-1} dp_s.$$

Using the Nosé-Poincaré method, we find the modified Nosé-Poincaré Hamiltonian $\tilde{H} = H + \epsilon H_1 + \mathcal{O}(\epsilon^2)$, such that $\epsilon = (\Delta t)^2$ and

$$\begin{aligned} H_1(\mathbf{q},s,\mathbf{p},p_s) &= \frac{1}{12}\left[\frac{p_s}{\mu}\mathbf{p}^TM^{-1}\nabla_{\mathbf{q}}U(\mathbf{q}) + \mathbf{p}^TM^{-1}U''(\mathbf{q})M^{-1}\mathbf{p} + \frac{2p_s^2}{\mu^2}gk_BT\right. \\ &- \left.\frac{1}{2}\nabla_{\mathbf{q}}U(\mathbf{q})^TM^{-1}\nabla_{\mathbf{q}}U(\mathbf{q}) - \frac{1}{2\mu}(\mathbf{p}^TM^{-1}\mathbf{p} - gk_BT)^2\right]. \end{aligned}$$

Following this,

$$\tilde{\rho} = \frac{1}{gk_BTC}\int e^{N_f\hat{\eta}}\frac{1}{|1+H_{\mathrm{N}}^0/(gk_BTe^{\hat{\eta}})|}\,dp_s.$$

We use Newton's method to solve $r(\eta) = 0$ with starting point

$$\eta_0 = -\frac{H + \frac{p_s^2}{2\mu} + h^2H_1 - H_{\mathrm{N}}^0}{gk_BT}$$

chosen so that $r(\eta_0) = -H_{\mathrm{N}}^0$. The first step gives

$$\eta_1 = \eta_0 + \frac{H_{\mathrm{N}}^0}{gk_BTe^{\eta_0} + H_{\mathrm{N}}^0}.$$

By construction, $r(\hat{\eta}) - r(\eta_1) = \mathcal{O}((H_{\mathrm{N}}^0)^2)$, and from a Taylor expansion, $r(\hat{\eta}) - r(\eta_1) = (\hat{\eta}-\eta_1)r'(\hat{\eta}) + \mathcal{O}((\hat{\eta}-\eta_1)^2)$. Therefore, $(\hat{\eta}-\eta_1)r'(\hat{\eta}) = \mathcal{O}((H_{\mathrm{N}}^0)^2)$. From the assumption that $gk_BTe^{\hat{\eta}} >> \mathcal{O}(H_{\mathrm{N}}^0)$, $r'(\hat{\eta}) = gk_BTe^{\hat{\eta}} + H_{\mathrm{N}}^0$ can be considered constant, and $\hat{\eta}-\eta_1 = \mathcal{O}((H_{\mathrm{N}}^0)^2)$. With the same assumption, we can approximate $\hat{\eta} = \eta_1$. Applying this to $\tilde{\rho}$ and expanding,

$$\begin{aligned} \tilde{\rho} &= \frac{1}{gk_BTC}\int e^{N_f\eta_0}\left(1 + \frac{(H_{\mathrm{N}}^0)}{gk_BTe^{\eta_0}}(N_f-1)\right)dp_s + \mathcal{O}(\Delta t^4) \\ &= \frac{\rho_c}{\bar{C}}e^{f_1(\mathbf{q},\mathbf{p})}\left(\int e^{f_2(p_s)}\,dp_s + \frac{H_{\mathrm{N}}^0(g-1)}{gk_BT}e^{\frac{-f_1(\mathbf{q},\mathbf{p})}{g} - \frac{H-H_{\mathrm{N}}^0}{gk_BT}}\int e^{(1-\frac{1}{g})f_2(p_s)}\,dp_s\right) + \mathcal{O}(\Delta t^4), \end{aligned}$$

where $\bar{C}$ is a constant, ρ_c is the canonical distribution,

$$f_1(\mathbf{q},\mathbf{p}) = \frac{-h^2}{24k_BT}\left(\sum_j\sum_k\frac{2p_jp_kU_{q_jq_k}}{m_jm_k} - \sum_j\frac{U_{q_j}^2}{m_j} - \frac{1}{\mu}\left(\sum_j\frac{p_j^2}{m_j} - gk_BT\right)^2\right),$$

and

$$f_2(p_s) = \frac{-1}{gk_BT}\left(\frac{p_s^2}{2\mu} + h^2\frac{1}{12}\left(\frac{p_s}{\mu}p^TM^{-1}\nabla_qU(q) + \frac{2p_s^2}{\mu^2}gk_BT\right)\right).$$

Because the ratio

$$\frac{\int e^{f_2(p_s)}\,dp_s}{\int e^{(1-\frac{1}{g})f_2(p_s)}\,dp_s} = \sqrt{1-\frac{1}{g}} + \mathcal{O}\left(\Delta t^4\right),$$

we can conclude that

$$\tilde{\rho} = \frac{\rho_c}{\tilde{C}}\exp\{f_1(\mathbf{q},\mathbf{p})\}\left(1 + \sqrt{1-\frac{1}{g}}\frac{H_{\mathrm{N}}^0}{k_BT}\exp\left\{\frac{-f_1(\mathbf{q},\mathbf{p})}{g} - \frac{H-H_{\mathrm{N}}^0}{gk_BT}\right\}\right) + \mathcal{O}\left(\Delta t^4\right),$$

where $\tilde{C}$ is a constant. □

Using this modified distribution $\tilde{\rho}$, we can derive

$$\omega_c = \rho_c/\tilde{\rho} + \mathcal{O}(\Delta t^2) \tag{8}$$

and the canonical average of A can be approximated by

$$\langle A \rangle_{c,\text{exact}} = \langle A\omega_c \rangle_{\text{num}} / \langle \omega_c \rangle_{\text{num}} + \mathcal{O}(\Delta t^4), \tag{9}$$

where $\langle \cdot \rangle_{\text{num}}$ denotes numerical average over a sufficiently long trajectory [5].

5. Practical computation

One drawback of computing more accurate averages is the complexity and cost involved in deriving the terms in the expansions presented in Section 4.1 or the weighting factor derived in Section 4.2. The modified vector field is typically derived using Taylor series and contains high-order derivatives of the vector field (or Hamiltonian). For systems arising in molecular simulation, this quickly becomes intractable beyond the first term in the series.

Skeel and Hardy have demonstrated that it is possible to compute a modified Hamiltonian to arbitrarily high-order without analytic calculation of higher derivatives [31]. This technique has been exploited to improve drastically the efficiency of hybrid Monte Carlo [1, 2, 19]. In this section, we apply similar finite differencing techniques to approximate many of the higher-order derivatives required for correcting averages.

5.1. Microcanonical ensemble with the Verlet method

For $\mathbf{w} = \nabla H$ and $\tilde{H}(\tilde{\mathbf{z}}) = 0$, a more accurate average can be calculated using (3). However, this adds additional computation at each timestep. The primary bottleneck is the $\mathcal{O}(N^2)$ calculation in the second derivative of H. Calculating ∇H is not a problem, since it is already required for timestepping. Computing $\nabla \cdot \nabla H$ does add some overhead, but it is an $\mathcal{O}(N)$ calculation, assuming ∇H can be computed in $\mathcal{O}(N)$ time.

By construction, $H(\hat{\mathbf{f}}) = \frac{1}{2}\left(\frac{H}{\nabla H \cdot \nabla H}\right)^2 \nabla H^T H'' \nabla H + \mathcal{O}(\epsilon^3)$, and thus the weight in (3) can be written as

$$\omega = \frac{\nabla \tilde{H} \cdot \nabla H}{\nabla H \cdot \nabla H} - H\frac{\nabla \cdot (\nabla H)}{\nabla H \cdot \nabla H} + 4\frac{H \circ \hat{\mathbf{f}}}{H} + \mathcal{O}(\epsilon^2). \tag{10}$$

This way, we do not need to calculate the second derivative of H.

Consider the modified Hamiltonian to be $\tilde{H} = H + \epsilon H_1 + \epsilon^2 H_2 + \mathcal{O}(\epsilon^3)$. We can choose the mapping function to be

$$\tilde{\mathbf{f}}(\mathbf{z}) = \mathbf{z} + \epsilon \frac{H_1(\mathbf{z})}{\nabla H(\mathbf{z}) \cdot \nabla H(\mathbf{z})} \nabla H(\mathbf{z}), \tag{11}$$

so that

$$H \circ \tilde{\mathbf{f}} = H + \frac{\epsilon H_1}{\nabla H \cdot \nabla H} \nabla H \cdot \nabla H + \frac{1}{2}\epsilon^2 H_1^2 \frac{\nabla H^T H'' \nabla H}{\nabla H \cdot \nabla H} + \mathcal{O}(\epsilon^3),$$

which means that

$$\begin{aligned} \frac{H \circ \tilde{\mathbf{f}}}{\epsilon\, H_1} &= \frac{H}{\epsilon H_1} + 1 + \frac{1}{2}\epsilon H_1 \frac{\nabla H^T H'' \nabla H}{\nabla H \cdot \nabla H} + \mathcal{O}(\epsilon^2) \\ &= \frac{-\epsilon^2 H_2}{\epsilon\, H_1} + \frac{1}{2}\epsilon H_1 \frac{\nabla H^T H'' \nabla H}{\nabla H \cdot \nabla H} + \mathcal{O}(\epsilon^2). \end{aligned}$$

Using all these approximations, we can conclude that

$$\omega = \frac{\nabla \tilde{H} \cdot \nabla H}{\nabla H \cdot \nabla H} + \epsilon H_1 \frac{\nabla \cdot (\nabla H)}{\nabla H \cdot \nabla H} - 4\frac{H \circ \tilde{\mathbf{f}}}{\epsilon H_1} - 4\epsilon \frac{H_2}{H_1} + \mathcal{O}(\epsilon^2).$$

Depending on the numerical method, H_1 and H_2 can be calculated by finite differencing in time, resulting in an $\mathcal{O}(N)$ calculation.

To calculate $\nabla\tilde{H}\cdot\nabla H$, we can use finite differencing as

$$\nabla\tilde{H}\cdot\nabla H=\left[\begin{array}{c}\frac{\partial\tilde{H}}{\partial q_i}\\ \frac{\partial\tilde{H}}{\partial p_i}\end{array}\right]\cdot\left[\begin{array}{c}\frac{\partial H}{\partial q_i}\\ \frac{\partial H}{\partial p_i}\end{array}\right]=\left[\begin{array}{c}-\frac{d\tilde{p_i}}{dt}\\ \frac{dq_i}{dt}\end{array}\right]\cdot\left[\begin{array}{c}\frac{\partial H}{\partial q_i}\\ \frac{\partial H}{\partial p_i}\end{array}\right]. \tag{12}$$

This way, we do not need to calculate $\tilde{H}$.

Let us choose the Verlet method

$$\left\{\begin{array}{lcl}\mathbf{p}^{n+1/2} &=& \mathbf{p}^n-\frac{1}{2}\Delta t\nabla U(\mathbf{q}^n)\\ \mathbf{q}^{n+1} &=& \mathbf{q}^n+\Delta t M^{-1}\mathbf{p}^{n+1/2}\\ \mathbf{p}^{n+1} &=& \mathbf{p}^{n+1/2}-\frac{1}{2}\Delta t\nabla U(\mathbf{q}^{n+1})\end{array}\right. \tag{13}$$

for our problem. The modified Hamiltonian $\tilde{H}$ of the Verlet method is $H+\epsilon H_1+\epsilon^2H_2+\mathcal{O}(\epsilon^3)$, where $\epsilon=\Delta t^2$ and

$$\begin{aligned}H_1 &= \frac{1}{24}(2M^{-1}\mathbf{p}U_{\mathbf{qq}}M^{-1}\mathbf{p}-\nabla_{\mathbf{q}}UM^{-1}\nabla_{\mathbf{q}}U),\\ H_2 &= \frac{1}{720}(-3(M^{-1}\nabla U)^TU^{''}(M^{-1}\nabla U)+12(M^{-1}p)^TU^{''}M^{-1}U^{''}(M^{-1}\mathbf{p})\\ &\quad +6U^{'''}\{M^{-1}\mathbf{p},M^{-1}\mathbf{p},M^{-1}\nabla U\}-U^{''''}\{M^{-1}\mathbf{p},M^{-1}\mathbf{p},M^{-1}\mathbf{p},M^{-1}\mathbf{p}\}).\end{aligned}$$

Converting spatial derivatives to time derivatives along the trajectory, this is the same as

$$\begin{aligned}H_1 &= \frac{1}{12}\frac{d}{dt}\sum_i\frac{p_i}{m_i}(\frac{\partial U}{\partial q_i})+\frac{1}{24}\sum_i\left(\frac{\partial U}{\partial q_i}\right)^2\frac{1}{m_i}-(\Delta t)^2\sum_i\frac{1}{72m_i}\left(\frac{d}{dt}\left(\frac{\partial U}{\partial q_i}\right)\right)^2+\mathcal{O}((\Delta t)^4),\\ H_2 &= \frac{1}{80}\sum_i\frac{1}{m_i}\left(\frac{d}{dt}\left(\frac{\partial U}{\partial q_i}\right)\right)^2-\frac{1}{1440}\frac{d^2}{dt^2}((\nabla U)^TM^{-1}\nabla U)-\frac{1}{720}\frac{d^3}{dt^3}(M^{-1}\mathbf{p}\cdot\nabla U)+\mathcal{O}((\Delta t)^2),\end{aligned}$$

which requires $\mathcal{O}(N)$ computation if computed using finite differencing in time.

The Verlet method calculates ∇H at each step, but $\nabla\tilde{H}$ must be calculated additionally for our weighting. Specifically,

$$\nabla H_1=\left[\begin{array}{c}\frac{\partial H_1}{\partial q_i}\\ \frac{\partial H_1}{\partial p_i}\end{array}\right]=\left[\begin{array}{c}\frac{1}{12}\left(\sum_{j,k}m_j^{-1}p_j\frac{\partial^3U}{\partial q_i\partial q_j\partial q_k}m_k^{-1}p_k-\sum_j\frac{\partial^2U}{\partial q_i\partial q_j}m_j^{-1}\frac{\partial U}{\partial q_j}\right)\\ \frac{1}{6}\left(\sum_j m_i^{-1}\frac{\partial^2U}{\partial q_i\partial q_j}m_j^{-1}p_j\right)\end{array}\right]. \tag{14}$$

Approximating ∇H_1 by differentiation with respect to time requires

$$\begin{aligned}\frac{d^2}{(dt)^2}(\frac{\partial U}{\partial q_i}) &= \sum_{j,k}m_j^{-1}p_j\frac{\partial^3U}{\partial q_i\partial q_j\partial q_k}m_k^{-1}p_k-\sum_j\frac{\partial^2U}{\partial q_i\partial q_j}m_j^{-1}\frac{\partial U}{\partial q_j}+\mathcal{O}((\Delta t)^2),\\ \frac{d^3q_i}{(dt)^3} &= \frac{1}{m_i}\frac{d}{dt}(\frac{\partial U}{\partial q_i})=\sum_j\left(m_i^{-1}\frac{\partial^2U}{\partial q_i\partial q_j}m_j^{-1}p_j\right)+\mathcal{O}((\Delta t)^2).\end{aligned}$$

Substituting these into (14),

$$\nabla H_1=\left[\begin{array}{c}\frac{1}{12}\frac{d^2}{(dt)^2}(\frac{\partial U}{\partial q_i})\\ \frac{1}{6m_i}\frac{d}{dt}(\frac{\partial U}{\partial q_i})\end{array}\right]+\mathcal{O}((\Delta t)^2).$$

This finite differencing scheme requires only a three-point stencil, rather than a five-point stencil as in the general case.

5.2. Canonical ensemble with the Nosé-Poincaré method

The equations of motion for the Nosé-Poincaré Hamiltonian are

$$\begin{cases} \dfrac{d\mathbf{q}}{dt} &= M^{-1}\dfrac{\tilde{\mathbf{p}}}{s} \\ \dfrac{d\mathbf{p}}{dt} &= -s\nabla_q U \\ \dfrac{ds}{dt} &= s\dfrac{p_s}{\mu} \\ \dfrac{dp_s}{dt} &= \dfrac{\tilde{\mathbf{p}}^T M^{-1}\tilde{\mathbf{p}}}{s^2} - gk_BT - \Delta H_{\mathrm{N}}. \end{cases}$$

Finite differencing in time can be used to compute the higher-order derivatives in the weighting factor. Since

$$\begin{aligned} \frac{d^2U}{dt^2} &= \frac{d}{dt}\left(\nabla_{\mathbf{q}}U\cdot\frac{d\mathbf{q}}{dt}\right) \\ &= \frac{d}{dt}\left(\nabla_{\mathbf{q}}U\cdot M^{-1}\frac{\tilde{\mathbf{p}}}{s}\right) \\ &= \left(M^{-1}\frac{\tilde{\mathbf{p}}}{s}\right)^T U''\left(M^{-1}\frac{\tilde{\mathbf{p}}}{s}\right) - \nabla_{\mathbf{q}}UM^{-1}\nabla_{\mathbf{q}}U - \nabla_{\mathbf{q}}U\cdot M^{-1}\tilde{\mathbf{p}}\frac{p_s}{s\mu}, \end{aligned}$$

we can conclude that

$$\mathbf{p}^T M^{-1}U''M^{-1}\mathbf{p} = \frac{d^2}{dt^2}U + \nabla_{\mathbf{q}}U\cdot M^{-1}\nabla_{\mathbf{q}}U + \frac{p_s}{s\mu}\nabla_{\mathbf{q}}U\cdot M^{-1}\tilde{\mathbf{p}}, \tag{15}$$

so that the weighting factor requires only $\mathcal{O}(N)$ computation, assuming that $\nabla_{\mathbf{q}}U$ is already computed at each timestep.

6. Numerical experiments

To test the methods derived in the previous sections, we use a system of 512 point particles, each with unit mass. We use a Lennard-Jones potential

$$U = 4\left(\frac{1}{r^{12}} - \frac{1}{r^6}\right),$$

where r is the distance between two particles. The particles are placed in a cubic box with periodic boundary conditions. The density is set to 0.95 and initial temperature is set to $k_BT = 1.5$. The timestep size is varied from 0.01 to 0.001. For the microcanonical problems we use the Verlet method (13) with the weighted average (3), choosing $\hat{\mathbf{f}}$ from (11) and the weight ω from (10). We use second-order central difference approximations to calculate (12) in the approximation of the weight (10). For the canonical problems we use the GLA (6) with the weighted average (9), choosing $\tilde{\rho}$ from (7) and the weight ω_c from (8). We use fourth-order central finite difference approximation to calculate (15) in the approximation of the weight (8).

In the problems considered here the exact solution is unknown, except for the temperature for the Nosé-Poincaré case. However, we can use the fact that the numerical results (should) converge to the exact value for a sufficiently long simulation time, as the timestep goes to zero. Since the unweighted numerical values have lower variance than the weighted ones, we use a sequence of unweighted values with differing timesteps to approximate the value when the timestep is zero. Of course, this extrapolation

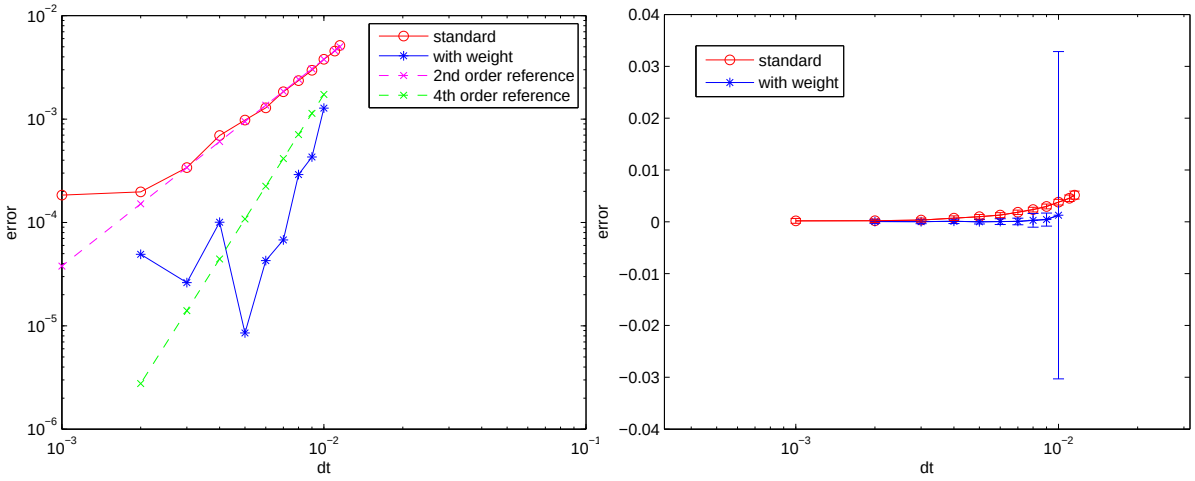

Figure 1: Error in internal energy as function of step size.

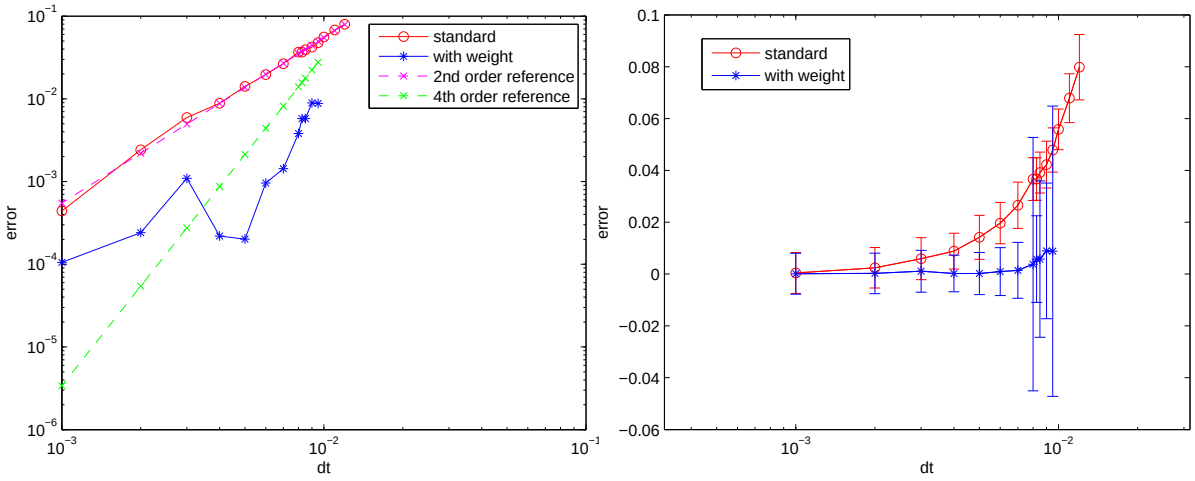

Figure 2: Error in scalar virial as function of step size.

procedure is not used by the proposed methodology, and is only used to help verify we are achieving the predicted higher-order of convergence.

Using a hypervirial theorem, the temperature can be calculated from

$$k_B T = \frac{\langle B \cdot \nabla H \rangle}{\langle \nabla \cdot B \rangle},$$

for any bounded C^1 vector field, B, which is periodic in $\mathbf{q}$ (see [20]). This result follows from

$$0 = \int \nabla \cdot (B e^{-\frac{1}{k_B T} H}) = \int \nabla \cdot B e^{-\frac{1}{k_B T} H} - \frac{1}{k_B T} \int B \cdot \nabla H e^{-\frac{1}{k_B T} H},$$

which implies

$$\int \nabla \cdot B e^{-\frac{1}{k_B T} H} = \frac{1}{k_B T} \int B \cdot \nabla H e^{-\frac{1}{k_B T} H},$$

assuming periodic boundary conditions, and that H is also periodic in $\mathbf{q}$. In our numerical experiments, we use $B = \mathbf{p}$ and $B = \nabla U$ for the temperature calculations.

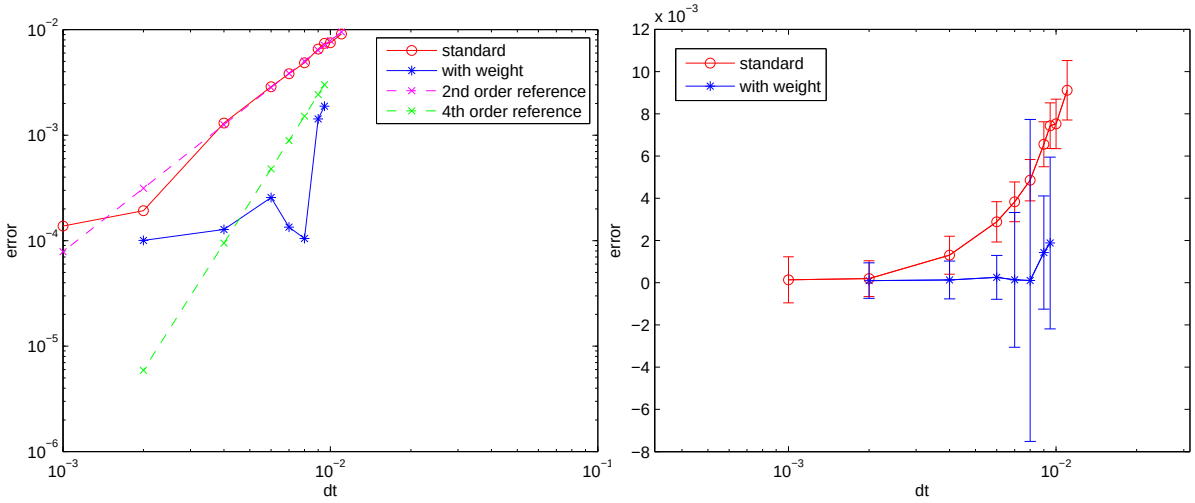

Figure 3: Error in kinetic energy temperature as function of step size.

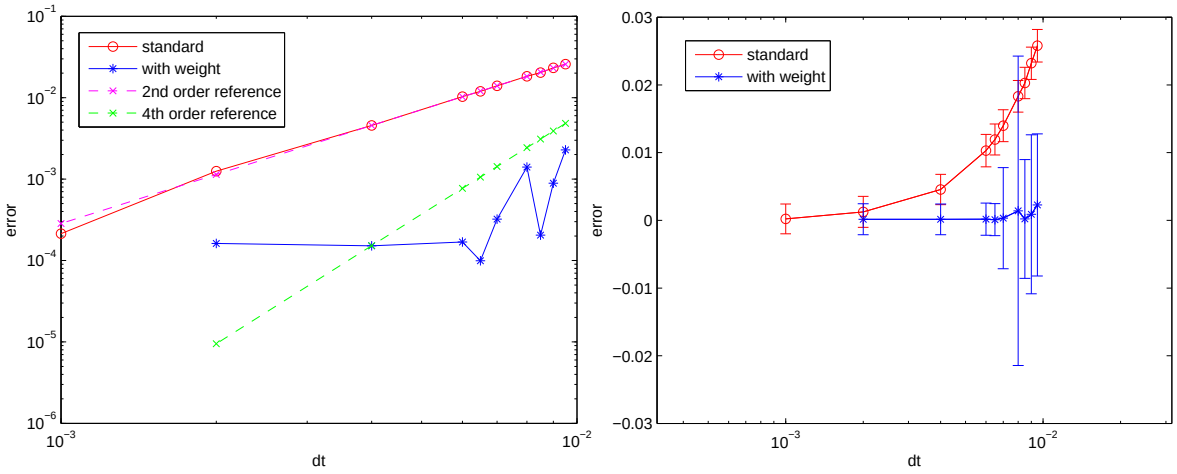

Figure 4: Error in hypervirial temperature as function of step size.

6.1. Microcanonical ensemble

We choose $\mathbf{w} = \nabla H$ and use a three-point finite difference approximation for ∇H_1 in these experiments. The numerical average of internal energy U, scalar virial $\mathbf{p}^T M\mathbf{q}$, temperature from kinetic energy $B = \mathbf{p}$, and temperature from hypervirial $B = \nabla U$ are plotted in Figs. 1, 2, 3, and 4, respectively. Average internal energy and temperature using hypervirial depend only on $\mathbf{q}$ and temperature using kinetic energy depends only on $\mathbf{p}$, but the scalar virial uses both $\mathbf{q}$ and $\mathbf{p}$. In addition, the approximation

$$H\frac{\nabla H H^{''} \nabla H}{(\nabla H \cdot \nabla H)^2} = 2\frac{H(\hat{f})}{H}$$

is used for the temperature using the hypervirial. All figures show averages from the Verlet method with and without weights, and second- and fourth-order reference lines. The statistical error bars were computed using the variance from over one hundred independent simulations, where each simulation uses one billion timesteps. The figures clearly show that the averages calculated with weights are better than

standard averages. However, we cannot conclude clear fourth-order convergence because of statistical error.

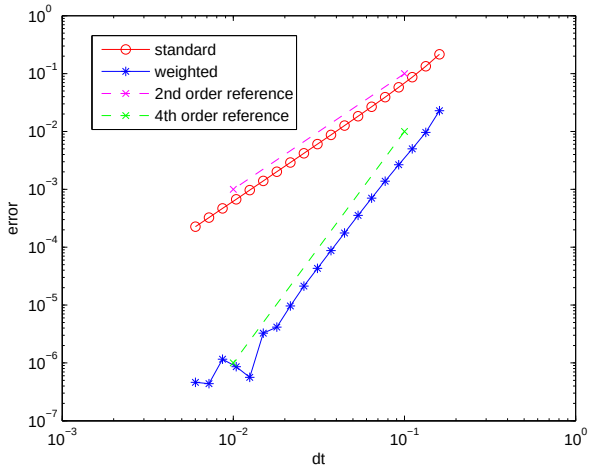

Figure 5: Error in average of $\sqrt{q_1^2 + q_2^2}$ as function of step size.

In the microcanonical ensemble, the temperature computed using different choices for B may differ by a term which is $\mathcal{O}(1/N)$ in magnitude [20]. Hence, we cannot make direct comparisons to determine a better approximate temperature. However, we can use differing estimates to help determine the statistical error in the numerical simulations.

6.2. Multiple invariants

Typical molecular dynamics simulations use periodic boundary conditions, which breaks conservation of angular momentum in N-body systems. To demonstrate that the methodology works for a system with angular momentum preservation, we apply the Verlet method to the modified Kepler problem

$$H = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 - \frac{1}{\sqrt{q_1^2 + q_2^2}} - \frac{0.015}{\sqrt{q_1^2 + q_2^2}^3},$$

and the mapping in Eq. (5). In this experiment, we calculate the average of $\sqrt{q_1^2 + q_2^2}$ with initial condition $(p_1, p_2, q_1, q_2) = (0, 2, 0.4, 0)$. Fig. 5 displays both standard and weighted averages. The figure clearly shows fourth-order convergence in the weighted average.

6.3. Canonical ensemble

To test the derived weighting methodology for canonical ensemble averages, we perform a sequence of numerical experiments using the Nosé-Poincaré method. In [5], it was suggested by the authors that the constant in the Nosé-Poincaré Hamiltonian could be perturbed so that the modified Nosé-Poincaré Hamiltonian was zero at $t = 0$, which allowed for significant simplification in the derivation of the weighting factor, ω_c. In Section 4.2, we generalized this result, computing a slightly different weighting factor which would allow for the use of an unperturbed constant, so that the unmodified Hamiltonian is zero at $t = 0$, as prescribed by original algorithm [6]. In the first experiment of this section, we compare both of these methods when applied to the task of computing the kinetic energy temperature. In the figures, "$H = 0$" corresponds to using an unperturbed value for the Nosé-Poincaré constant and "Hmodified $= 0$" corresponds to using a perturbed value, as suggested in [5]. The results are shown in Fig. 6. Although the use of a perturbed constant is more accurate, both weighting methods result in higher-order averages. In a second experiment, we use an unperturbed Nosé-Poincaré constant, and calculate the hypervirial temperature, which depends on $\mathbf{q}$ instead of $\mathbf{p}$. This also results in the expected convergence, as shown in Fig. 7. In a third experiment, we calculate the scalar virial, which depends on both $\mathbf{q}$ and $\mathbf{p}$, again using an unperturbed constant. The results in Fig. 8 show fourth-order convergence for the weighted average error.

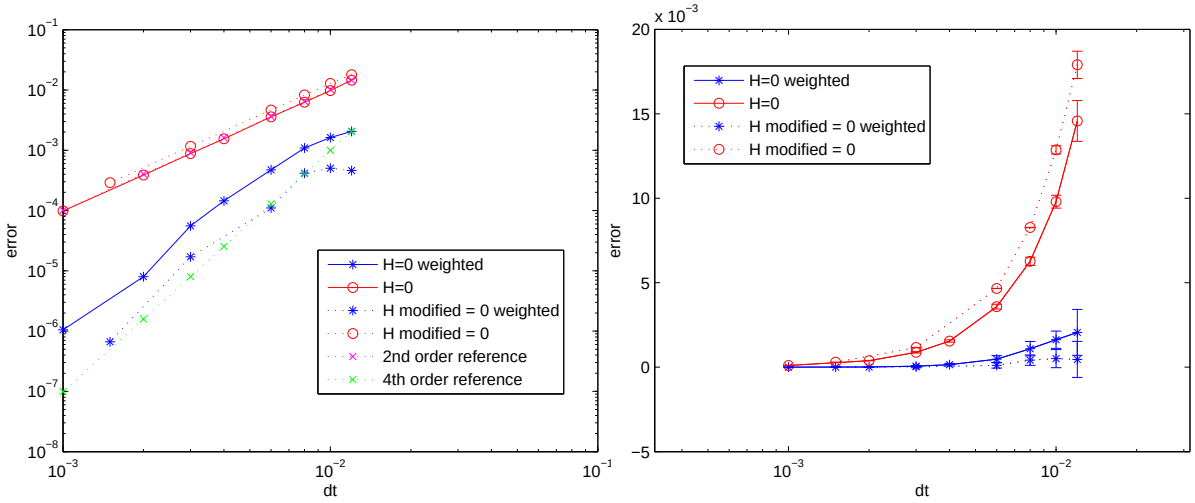

Figure 6: Error in kinetic energy temperature as function of step size.

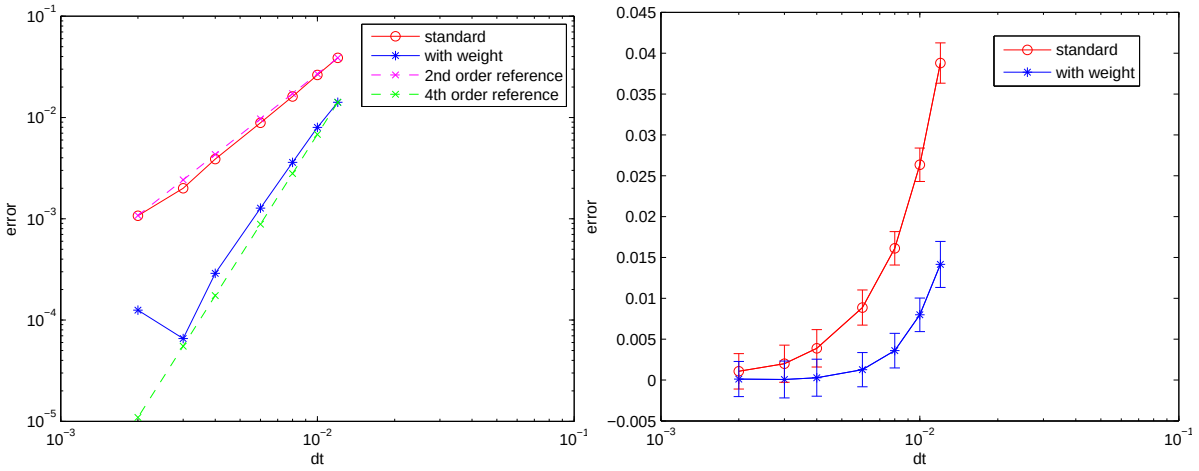

Figure 7: Error in hypervirial temperature as function of step size.

6.4. Computational cost

To measure the computational overhead associated with computing more accurate averages, we measure the simulation time for a short simulation for computing the kinetic energy temperature, using 10,000 steps with a timestep size 0.01. In our simple implementation, the cost of a single force evaluation scales quadratically with the number of particles (we are not using neighbor lists to reduce the complexity). Fig. 9 shows the actual simulation time in seconds as a function of the number of particles. The left panel shows the simulation time for computing weighted and unweighted microcanonical averages, using the Verlet method. Computing the (more accurate) weighted averages increases the computational cost as the number of particles grows, however the overall complexity is still quadratic in the number of particles (the same as the complexity of a force evaluation). The right panel shows the simulation time for computing weighted and unweighted canonical averages, using the Nosé-Poincaré method. The top curve corresponds to the simulation time for computing (the more accurate) weighted averages, but without using any finite difference approximations. The two middle curves correspond to simulations where we compute (a) standard unweighted averages and (b) weighted averages but using finite difference approxi-

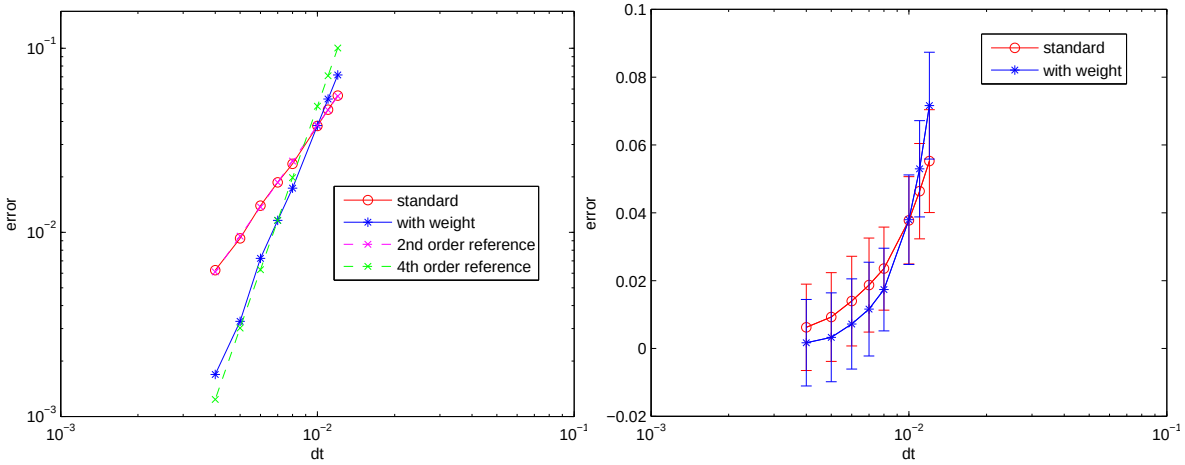

Figure 8: Error in scalar virial as function of step size.

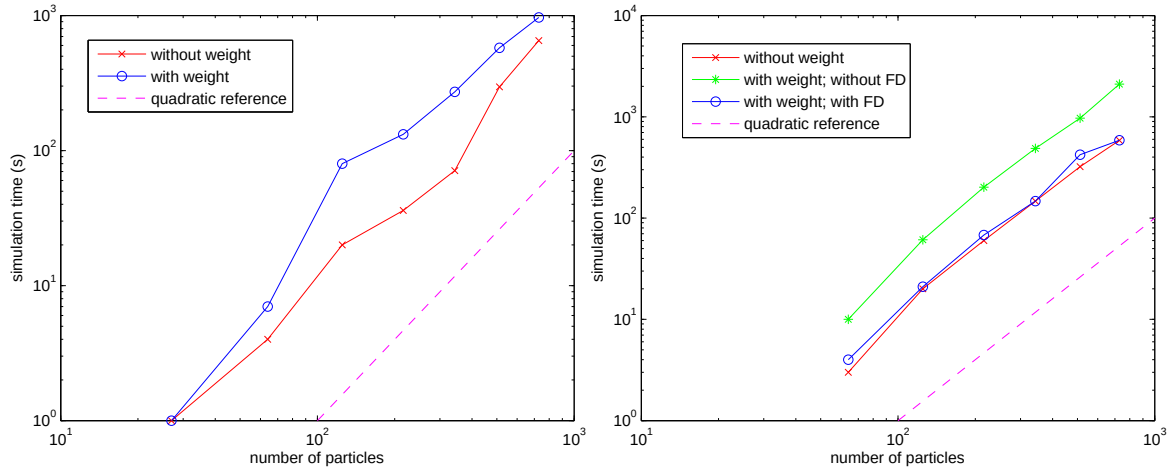

Figure 9: Simulation time as a function of the number of particles. The left panel shows the simulation time for computing weighted and unweighted microcanonical averages. The right panel shows the simulation time for computing weighted and unweighted canonical averages.

mations. In this case the additional cost of computing the weight is insignificant, so long as we use finite differencing to reduce the cost of computing the weight.

7. Conclusion

Given a Hamiltonian system and a symplectic integrator, we have derived new weighting methods to estimate and correct the truncation error in molecular dynamics averages. We showed that these new methods can be used to reduce the numerical error and enhance the order of convergence without additional high-overhead calculations. It is possible to extend this approach to any symplectic integrator applied to a Hamiltonian system with known first integrals.

Acknowledgments

The authors wish to thank Prof. Michael Heath for helping improve the written presentation. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

[1] E. Akhmatskaya, S. Reich, GSHMC: An efficient method for molecular simulation, J. Comput. Phys. 227 (2008) 4934–4954.

[2] E. Akhmatskaya, S. Reich, New hybrid Monte Carlo methods for efficient sampling: from physics to biology and statistics, Prog. Nucl. Sci. Tech. 2 (2011) 447–462.

[3] M. P. Allen, D. J. Tildesley, Computer Simulation of Liquids, Oxford Science, Oxford, 1987.

[4] G. Benettin, A. Giorgilli, On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms, J. Stat. Phys. 74 (1994) 1117–1143.

[5] S. D. Bond, B. J. Leimkuhler, Molecular dynamics and the accuracy of numerically computed averages, Acta Numerica 16 (2007) 1–65.

[6] S. D. Bond, B. J. Leimkuhler, B. B. Laird, The Nosé-Poincaré method for constant temperature molecular dynamics, J. Comput. Phys. 151 (1) (1999) 114–134.

[7] E. Cancés, F. Castella, P. Chartier, E. Faou, C. Le Bris, F. Legoll, G. Turinici, High-order averaging schemes with error bounds for thermodynamical properties calculations by molecular dynamics simulations, J. Chem. Phys. 121 (21) (2004) 10346–10355.

[8] E. Cancés, F. Castella, P. Chartier, E. Faou, C. Le Bris, F. Legoll, G. Turinici, Long-time averaging for integrable Hamiltonian dynamics, Numer. Math. 100 (2) (2005) 211–232.

[9] D. Chandler, Introduction to Modern Statistical Mechanics, Oxford University Press, New York, 1987.

[10] M. A. Cuendet, W. F. van Gunsteren, On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm, J. Chem. Phys. 127 (18) (2007) 184102.

[11] R. L. Davidchack, Discretization errors in molecular dynamics simulations with deterministic and stochastic thermostats, J. Comput. Phys. 229 (24) (2010) 9323–9346.

[12] C. P. Dettmann, G. P. Morriss, Hamiltonian reformulation and pairing of Lyapunov exponents for Nosé-Hoover dynamics, Phys. Rev. E 55 (3) (1997) 3693–3696.

[13] M. P. Eastwood, K. A. Stafford, R. A. Lippert, M. Ø. Jensen, P. Maragakis, C. Predescu, R. O. Dror, D. E. Shaw, Equipartition and the calculation of temperature in biomolecular simulations, J. Chem. Theory Comput. 6 (7) (2010) 2045–2058.

[14] D. Frenkel, B. Smit, Understanding Molecular Simulation, 2nd ed., Academic Press, New York, 2002.

[15] E. Hairer, Backward analysis of numerical integrators and symplectic methods, Annals of Numerical Mathematics 1 (1994) 107–132.

[16] E. Hairer, C. Lubich, The lifespan of backward error analysis for numerical integrators, Numer. Math. 76 (1997) 441–462.

[17] E. Hairer, C. Lubich, G. Wanner, Geometric numerical integration illustrated by the Störmer–Verlet method, Acta Numerica 12 (2003) 399–450.

[18] J. P. Hansen, I. R. McDonald, Theory of Simple Liquids, 2nd ed., Academic Press, New York, 1986.

[19] J. Izaguirre, S. Hampton, Shadow hybrid Monte Carlo: An efficient propagator in phase space of macromolecules, J. Comput. Phys. 200 (2) (2004) 581–604.

[20] O. G. Jepps, G. Ayton, D. J. Evans, Microscopic expressions for the thermodynamic temperature, Phys. Rev. E 62 (4) (2000) 4757–4763.

[21] B. J. Leimkuhler, S. Reich, R. D. Skeel, Integration methods for molecular dynamics, in: IMA Volumes in Mathematics and its Applications, vol. 82, Springer-Verlag, New York, 1996, pp. 161–186.

[22] D. A. McQuarrie, Statistical Mechanics, Harper and Row, New York, 1976.

[23] A. I. Neishtadt, The separation of motions in systems with rapidly rotating phase, J. Math. Mech. 48 (2) (1984) 133–139.

[24] S. Nosé, A molecular-dynamics method for simulations in the canonical ensemble, Mol. Phys. 52 (1984) 255–268.

[25] S. Nosé, A unified formulation of the constant temperature molecular-dynamics methods, J. Chem. Phys. 81 (1) (1984) 511–519.

[26] S. Reich, Backward error analysis for numerical integrators, SIAM J. Numer. Anal. 36 (1999) 1549–1570.

[27] J. M. Sanz-Serna, Geometric integration, in: I. S. Duff, G. A. Watson (eds.), The State of the Art in Numerical Analysis, Clarendon Press, Oxford, 1997, pp. 121–143.

[28] J. M. Sanz-Serna, M. P. Calvo, Numerical Hamiltonian Problems, Chapman and Hall, New York, 1994.

[29] T. Schlick, Molecular Modeling and Simulation, Springer-Verlag, New York, 2002.

[30] R. D. Skeel, What makes molecular dynamics work?, SIAM J. Sci. Comput. 31 (2009) 1363–1378.

[31] R. D. Skeel, D. J. Hardy, Practical construction of modified Hamiltonians, SIAM J. Sci. Comput. 23 (4) (2001) 1172–1188.

[32] M. Toda, R. Kubo, N. Saitô, Statistical Physics I, 2nd ed., Springer-Verlag, New York, 1991.

[33] P. F. Tupper, Computing statistics for Hamiltonian systems: A case study, J. Comput. Appl. Math. 205 (2) (2007) 826–834.

[34] H. Yoshida, Non-existence of the modified first integral by symplectic integration methods, Phys. Lett. A 282 (4–5) (2001) 276–283.